CHAPTER 5

Databases, Sampling Design, and Data Analysis

Developing National/Regional Databases Sampling Design Quality Assurance/Quality Control Statistical Analyses

5.1 INTRODUCTION

Development of national regional numeric nutrient criteria requires that an extensive amount of data from across the country be evaluated. This information can be an invaluable tool to States and Tribes as they develop nutrient criteria. Both existing and historical data may provide considerable information that is specific to the region where criteria are to be set. First the data must be located, then the suitability of the data (type and quality) ascertained before they can be used for analysis of water quality parameters. It is also important to determine how the data were collected to make future monitoring efforts compatible with earlier approaches. Descriptive data that characterize the waterbody are invaluable.

Data may come from existing sources or can be collected from new sampling programs. Nutrient-related data for estuaries and coastal waters, collected by various agencies for many different purposes, exist in numerous databases and have the potential to provide the basis for development of nutrient criteria on a regional level. This chapter presents an overview of existing databases and a general discussion concerning the evaluation of such datasets in terms of their use in the nutrient criteria development process. The list of databases is not all-inclusive—many other data sources exist—but the list provided is intended to represent the kind of information that is available. This chapter also provides a description of existing data resources (e.g., U.S. EPA Legacy STORET and ODES) and how these data may be used to generate preliminary nutrient criteria on regional levels. In addition to discussing the use of existing data, the chapter discusses new data collection, including consideration for sampling design and the types of sampling to be considered as part of data collection activities. The chapter ends with a general discussion of data management, quality assurance, and quality control issues that are integral in the overall discussion of data storage, accessibility, and utilization.

5.2 DEVELOPING REGIONAL AND NATIONAL DATABASES FOR ESTUARIES AND COASTAL WATERS

A database is a collection of information related to a particular subject or purpose. Databases are arranged so that they divide data into separate electronic repositories in tabular format. Data in tables can be viewed and edited, and new data can be added. A single datum is stored in only one table but can be viewed from multiple locations. Updating one view of a datum will update it in all the various viewable forms. Each table should contain a specific type of information. Data from different tables can be viewed simultaneously according to the user-defined table relationships. That is, the relationship among data in different tables can be defined so that more than one table can be queried or reported and accessed in a single view. Data stored in tables can be located and retrieved using queries. A query allows the user to find and retrieve only the data that meet user-specified conditions. Queries also can be used to update or delete multiple records simultaneously and to perform built-in or custom calculations of data. Data in tables can be analyzed and printed in specific layouts for reports.

To facilitate data manipulation and calculations, it is highly recommended that historical and present-day data be transferred to a relational database. A relational database is a collection of data items organized as a set of formally described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. Each table contains one or more data categories in columns. Each row contains a unique instance of data for the categories defined by the columns. The organization of data into relational tables is known as the logical view of the database. Relational databases are powerful tools for data manipulation and initial data reduction. They allow selection of data by specific and multiple criteria and definition and redefinition of linkages among data components.

Geographic information systems (GIS) are geo-referenced databases that have a geographic component (i.e., spatial platform) in the user interface. Spatial platforms associated with a database allow geographic display of sets of sorted data and make mapping easier. These types of databases with spatial platforms are becoming more common. The system is based on the premises that “a picture is worth a thousand words” and that most data can be related to a map or other easily understood graphic. GIS platforms such as ArcView, ArcInfo, and MapInfo are frequently used to integrate spatial data with monitoring data for watershed analysis.

The EPA National Nutrient Criteria Program initiated the development of a national database application that will be used to store and analyze nutrient data. The ultimate use of these data will be to derive ecoregion- and waterbody-type specific numeric nutrient criteria. Initially, EPA developed a Microsoft Access application that was populated with STORET Legacy data, U.S. Geological Survey (USGS) (NAWQA, NASQAN, and Benchmark) data, and other relevant nutrient data from universities, States/Tribes, and additional data-rich entities. To serve the general public more effectively and efficiently, EPA also developed and maintains a web-accessible nutrient database application in an Oracle™ environment that allows for easy web accessibility, geo-referencing/GIS compatibility, and data analysis on a State/Tribal, regional, and national basis. The total amount of existing nutrient data nationally is large (>20 gigabytes), and it is anticipated that more data will be entered into the system. The Oracle™ application can easily manage large quantities of data and provides ample room for expansion as more data are collected. The Oracle™ database application is being designed for compatibility with EPA’s modernized STORET. A key feature of the database design will prevent duplication of effort for users of STORET and the nutrients database application, especially for data updating. Considerable efforts are also being made to ensure compatibility with other database systems (e.g., WQS and RAD) currently being developed by EPA’s Office of Water. The Oracle™ application has been online since the fall of 2000.

Data Sources

Potential sources of data include water quality monitoring data from Federal, State, Tribal, and local water quality agencies; university studies; and volunteer monitoring programs. However, the data sources described in this section do not encompass the full extent of available data sources. The data available in the nutrient database can be used to identify reference areas to begin development of potential nutrient criteria. The nutrient data sources for estuaries and coastal waters that will be useful for developing criteria are discussed below. These data sources contain extensive water quality data, however, data; collection should not be limited to these sources. Collection of scientifically sound water quality data from any reliable source is encouraged.

Many of the water quality programs listed here include rivers and streams data or mixed freshwater, estuarine, and coastal water systems data. The rivers and streams information is included in this document because it gives relevant data about nutrient loading from fluvial systems, which is important to estuaries and coastal waters. Generally, in estuaries that have been impaired by nutrients, a database exists, and in less impaired estuaries, the database is often insufficient for comparisons. Nutrient loading information from fluvial systems may provide a basis for comparison between systems if they share important geophysical conditions. Such comparisons would assist in developing trends and extrapolating where insufficient data exist.

EPA Water Quality Data

EPA has many programs of national scope that focus on collection and analysis of water quality data. The following information on several of the databases and national programs may be useful to water quality managers as they compile data for criteria development.

STOrage and RETrieval System (STORET)

STORET is EPA's national database for water quality and biological data. EPA's original STORET system, called the STORET Legacy Data Center (LDC) and operated continuously since the 1960s, was historically the largest repository of water quality data in the Nation. This legacy mainframe-based system was the repository of all data held in EPA's original STORET system as of the end of 1998. This Legacy STORET ceased to exist in the year 2000. In its place, EPA is supporting a modernized database, simply called STORET, designed as a replacement for the original STORET System. While STORET will serve as the major repository for more current data, the nutrient criteria database application will offer major improvements in database content and capabilities that will enable more detailed data analysis.

Interested parties may view both databases on the World Wide Web. For the nutrient database, capabilities exist to produce printed reports and download data files. Queries for data via the web will be designed for use by the general public and will require no special training or software.

STORET is a compendium of data supplied by Federal, State, and local organizations used to evaluate environmental conditions in the field. The data in STORET are organized by both geographic location and data ownership. Every field study site is identified by at least one latitude/longitude and, where appropriate, also by State/Province, county, drainage basin, and stream reach. Monitoring activities recorded include field measurements, habitat assessments, water and sediment samples, and biological population surveys. Records cover the complete spectrum of physical properties, concentrations of substances, and abundance and distribution of species observed during biological monitoring. STORET is designed for maximum compatibility with commercial software, including GISs such as the ESRI ArcView package, and statistical packages such as PC SAS. STORET downloaded files import easily into all standard spreadsheet packages. Further information about STORET may be obtained by e-mailing STORET@epa.gov, or telephoning toll-free at 1-800-424-9067.

Environmental Monitoring and Assessment Program (EMAP)

EMAP is an EPA research program designed to develop the tools necessary to monitor and assess the status and trends of national ecological resources (see EMAP Research Strategy on the EMAP website: www.epa.gov/emap). EMAP's goal is to develop the scientific understanding for translating environmental monitoring data from multiple spatial and temporal scales into assessments of ecological condition and forecasts of future risks to the sustainability of the Nation's natural resources. EMAP's research supports the National Environmental Monitoring Initiative of the Committee on Environment and Natural Resources (CENR). Data from EMAP can be downloaded directly from the EMAP website. The EMAP Data Directory contains information on available datasets, including data and metadata (language that describes the nature and content of data). The status of the Data Directory as well as composite data and metadata files also are available on the EMAP website. EMAP–estuaries data is one of several areas addressed by the program. Most of the estuaries data were collected during a summer index period.

Ecological Data Application System (EDAS)

EDAS is EPA's program-specific counterpart to STORET. EDAS was developed by EPA's Office of Water to manipulate data obtained from biological monitoring and assessment and to assist States/Tribes in developing biocriteria. It contains built-in data reduction and recalculation queries that are used in biological assessment. The EDAS database is designed to enable the user to easily manage, aggregate, integrate, and analyze data to make informed decisions regarding the condition of a water resource. Biological assessment and monitoring programs require aggregation of raw biological data (lists and enumeration of taxa in a sample) into informative indicators. EDAS is designed to facilitate data analysis, particularly the calculation of biological metrics and indexes. Predesigned queries that calculate a wide selection of biological metrics are included with EDAS. Future versions of EDAS will include the capability to upload data to, and download data from, the distributed version of modernized STORET. EDAS is not a final data warehouse, but it is a program or project-specific customized data application for manipulating and processing data to meet user requirements. The EDAS application is currently under development; more information will be available through the EPA website.

Ocean Data Evaluation System (ODES)

ODES is used for storing and analyzing water quality and biological data from marine, estuarine, and some freshwater environments. The system supports Federal, State, and local decisionmakers associated with marine monitoring programs and managers and analysts who must meet regulatory objectives through the evaluation of marine monitoring information. ODES contains data from the National Estuary Program, the Great Lakes National Program Office, the Ocean Disposal Program, the 301(h) Sewage Discharge Program, the National Pollutant Discharge Elimination System Program, and the 403(c) Program. Records pertain to water quality, fish abundance, bioaccumulation, benthic infauna, fish histopathology, bioassay, and sediment physical/chemical characteristics. Users can examine both spatial and temporal relationships among variables. A quality assurance report describing analytical methods and procedures for each dataset is stored with each dataset.

Chesapeake Bay Program (CBP)

CBP, a cooperative effort between the Federal Government, the States, the District of Columbia, and local governments in the Chesapeake Bay watershed, provides funds to the States of Maryland and Virginia for the routine monitoring of 19 directly measured water quality parameters at 49 stations within the bay watershed. The Water Quality Monitoring Program began in June 1984 with stations sampled once each month during the colder late fall and winter months and twice each month in the warmer months. A refinement in 1995 reduced the number of monitoring cruises to 14 per year. Data are available on the internet at www.chesapeakebay.net/data/.

National Estuarine Programs (NEPs)

Many NEPs have nutrient and related data that could be used for characterization purposes. Presently, there is no national repository of NEP data, but, in the development of regional nutrient criteria, the NEPs may serve as an excellent source for information. Some of these programs have electronic databases and some hard copy data that could be acquired. EPA is attempting to acquire the available NEP data and eventually enter them into the National Nutrient Criteria Program database. A list of NEP estuarine systems can be found online at www.epa.gov/nep.

National Oceanographic and Atmospheric Administration (NOAA)

Water Quality Data in the National Oceanographic Data Center (NODC)

NODC is one of three national environmental data centers operated by NOAA and serves as a national repository and dissemination facility for global oceanographic data. Its primary mission is to ensure that global oceanographic data collected at great cost are maintained in a permanent archive easily accessible to the world science community and to other users. NODC holds physical, chemical, and biological oceanographic data collected by U.S. Federal agencies, including the Department of Defense (primarily the U.S. Navy); State and local government agencies; universities and research institutions; and private industry. NODC does not conduct any data collection programs of its own; it serves solely as a repository and dissemination facility for data collected by others (see website at www.nodc.noaa.gov). NODC provides data management support for major ocean science projects such as Tropical Ocean–Global Atmosphere (TOGA), World Ocean Circulation Experiment (WOCE), and Joint Global Ocean Flux Study (JGOFS). NODC's global holdings of physical, chemical, and biological oceanographic data include substantial amounts of data from coastal ocean areas. For example, the NODC Oceanographic Profile Database holds primarily coastal data (www.nodc.noaa.gov/cgi-bin/JOPI/jopi).

National Estuarine Research Reserve System (NERR)

The NERR Systemwide Monitoring Program was designed to identify and track short-term variability and long-term changes in the integrity and biodiversity of representative estuarine ecosystems and coastal watersheds for the purposes of contributing to effective national, regional, and site-specific coastal zone management. The program has two major goals: (1) to support State-specific nonpoint source pollution control programs by establishing local networks of continuous water quality monitoring stations in representative protected estuarine ecosystems; and (2) to develop a nationwide database on baseline environmental conditions in the NERR system of estuaries. Water quality data collected from phase 1 of the NERR Systemwide Monitoring Program provides data necessary for site and intersite baseline studies, trend analysis, and impact assessment. Data are available for each of the participating NERR systems at http://inlet.geol.sc.edu/cbmoweb/30_minute_data.html.

Rivers and Streams Water Quality Data

Rivers and streams water quality data are potentially useful for estuaries and coastal waters. Because much of the nutrient load to estuaries comes from rivers and streams, it is critical to define nutrient concentrations landward of tidal influence and to calculate fluvial-based nutrient loads to estuaries and potentially to coastal waters. EPA STORET, which was discussed in detail previously, includes data from rivers and streams from across the Nation. Another comprehensive Federal source of river and stream water quality data is USGS. USGS maintains databases on water quantity and quality for waterbodies across the Nation. Many of the data for rivers and streams are available through the National Water Information System (NWIS). The most convenient method of accessing the local databases is through the USGS State representative. Every State office can be reached through the USGS home page on the Internet at URL http://www.usgs.gov/wrd002.html. The USGS data from several national water quality programs covering large regions offer highly controlled and consistently collected data that may be particularly useful for nutrient criteria analysis. Two programs, the Hydrologic Benchmark Network (HBN) and the National Stream Quality Accounting Network (NASQAN), include routine monitoring of rivers and streams during the past 30 years. The USGS National Water Quality Assessment (NAWQA) Program is building a third national database of stream quality data collected and analyzed for more than 50 river basins and aquifer systems across the Nation. More information and data from each of these studies can be found on the USGS website (www.usgs.gov). For additional data sources, the Rivers and Streams Nutrient Criteria document presents an extensive list of related freshwater nutrient-related information (Nutrient Criteria Technical Guidance Manual—Rivers and Streams, 2000, EPA-822-B-00-002).

USGS San Francisco Bay Program

Since 1968, USGS has sustained a research program to understand how coastal ecosystems function and how those functions are altered by human disturbances. One component of this program is directed to following and understanding changes in the water quality of San Francisco Bay. The program includes regular measurements of water quality along a 145-kilometer transect spanning the length of the entire estuarine system, from the South Bay to the Sacramento River. The program studies many different aspects of San Francisco Bay, such as changing land use, hydrology, water currents, nutrients, toxic contaminants, geological structure, and biological communities. The results of water quality measurements, and eventually the full dataset, can be accessed at http://sfbay.wr.usgs.gov/access/index/wqdata.html.

State/Tribal Monitoring Programs

Most States monitor some estuaries and coastal waters within their borders for algal and nutrient variables. Data collected by State/Tribal water quality monitoring programs can be used for nutrient criteria development. These data should be available from the agencies responsible for monitoring.

Sanitation Districts

Massachusetts Water Resources Authority (MWRA)

MWRA has conducted a comprehensive monitoring program in Boston Harbor, Massachusetts Bay, and Cape Cod Bay from 1992 to the present. The program was established to understand baseline conditions and monitor the effects of effluent discharges into Boston Harbor and Massachusetts Bay. This multifaceted monitoring program focuses on water quality, benthic ecosystem health, effluent characterization, and public health issues related to metal, organic, and microbiological contaminants. All data are stored in an Oracle™ Relational Database Management System to support the monitoring program.

New York City–Department of Environmental Protection (NYCDEP)

NYCDEP has conducted extensive monitoring to evaluate viable treatment options for sewage effluents to mitigate conditions that promote eutrophication. The city's monitoring programs have included point source, water column, sediment, hydrodynamic and atmospheric studies. All data are stored in the NYCDEP databases and have been used in the development and application of a System-Wide Eutrophication Model (SWEM) to enhance the city's ability to evaluate the effectiveness of various treatment options in mitigating conditions that promote eutrophication.

Southern California

The major southern California dischargers of treated sewage effluents into marine waters have conducted applied research and monitoring programs for more than 30 years. These dischargers include the cities of San Diego and Los Angeles and the counties of Ventura/Oxnard, Los Angeles, Orange, and San Diego. The programs are designed to monitor the concentrations and mass emission rates of effluent materials in the treated effluent; the transport and fate of these materials in the receiving waters; the exposure of the contaminants to organisms in the receiving waters; and the effects of that exposure to individuals, populations, and communities of subtidal, intertidal, and water column organisms. Some of this monitoring is performed to comply with NPDES monitoring efforts, and other monitoring addresses specific issues of interest to the districts. The data are retained on a number of local databases, but they are also maintained on EPA's ODES. In addition to the localized databases managed by the sanitation districts, a research organization (Southern California Coastal Water Research Project) has performed parallel and specialized monitoring and applied research on the effects of treated sewage effluents in this region since the early 1970s. Their data are managed onsite and are provided to national data inventories (e.g., ODES, NODC, STORET).

California and Oregon

Similar monitoring requirements are established for other locations in California and Oregon. The districts that have effluents that are discharged into rivers, streams, or oceans are required, through the NPDES permits, to monitor their treated effluents and the receiving waters. The data are retained locally, but must also be filed with ODES. The State and Regional Water Resources Control Boards of California and the Departments of Ecology in Washington and Oregon administer these permits.

Puget Sound, Washington

The cities and counties on the Puget Sound watershed monitor the treated sewage effluents and the receiving waters in compliance with State water quality parameters. These data are provided to the State in electronic format and are retained on database systems administered by the Department of Ecology that are available to ODES. The data in the receiving water environments are collected in methods that are historically similar to work that has been performed in Puget Sound since the middle 1960s.

Academic and Literature Sources

Many research studies conducted by academic institutions may provide data useful for developing nutrient criteria. Academic research tends to be site specific and span a limited number of years, although data for some systems may span 20 years or more. Academic research data should be available from researchers. However, the scientific literature is likely to be a major source of estuarine and coastal waters data.

Volunteer Monitoring Programs

Many States have volunteer water quality monitoring programs. Some programs are State sponsored, while others are independently organized. Citizens in many areas donate their time, money, or experience to aid State, Tribal, and local governments in collecting water quality data. Volunteers analyze water samples for dissolved oxygen, nutrients, pH, temperature, and a host of other water constituents; evaluate the health of stream habitats and aquatic biological communities; note shore zone conditions and land uses that may affect water quality; catalogue and collect beach debris; and restore degraded habitats.

State and local agencies may use volunteer data to screen for water quality problems, establish trends in waters that would otherwise be unmonitored, and make planning decisions. Volunteers benefit from learning more about their local water resources and identifying what conditions or activities might contribute to pollution problems. As a result, volunteers frequently work with clubs, environmental groups, and State/Tribal or local governments to gather information and address problem areas. As with any other data source, whether student, State, Federal, academic, or volunteer based, documented quality assurance procedures are an important consideration.

EPA supports volunteer monitoring and local involvement in protecting our water resources. EPA support takes many forms, including sponsoring national and regional conferences to encourage information exchange among volunteer groups, government agencies, businesses, and educators; publishing sampling methods manuals for volunteers; and providing technical assistance (primarily on quality control and laboratory methods) and regional coordination through the 10 EPA Regional Offices. EPA also produces a Nationwide Directory of Volunteer Monitoring Programs, which is available online at http://yosemite.epa.gov/water/volmon.nsf. This directory lists volunteer organizations around the country engaged in monitoring rivers, lakes, estuaries, beaches, wetlands, and groundwater, as well as surrounding lands. EPA volunteer monitoring activities are coordinated in part through a website that lists many resources at http://www.epa.gov/owow/monitoring/volunteer.

Quality of Historical Data

The value of older historical datasets is a recurrent problem because data quality is often unknown. Knowledge of data quality is also problematic for long-term data repositories such as STORET and long-term State databases, where objectives, methods, and investigators may have changed many times over the years. The most reliable data tend to be those collected by a single agency, using the same protocol, for a limited number of years. Supporting documentation should be examined to determine the consistency of sampling and analysis protocols. Investigators must determine the acceptability of data contained in large, heterogeneous data repositories. Considerations and requirements for acceptance of these data are described below.

Location Data

STORET and USGS data are geo-referenced with latitude, longitude, and up to Reach File 3 (RF1 & RF3) codes. Geo-reference data can be used to select specific locations or specific USGS hydrologic units. In addition, STORET often contains a site description. Knowledge of the rationale and methods of site selection from the original investigators may supply valuable information. Metadata of this type, when known, are frequently stored within large long-term databases.

Variables and Analytical Methods

Thousands of variables are recorded in database records. Each separate analytical method yields a unique variable. For example, five ways of measuring total nitrogen (TN) results in five unique variables. We do not recommend mixing analytical methods in sample analyses because methods differ in accuracy, precision, and detection limits. Sample analyses should concentrate on a single analytical method for each parameter of interest. Selection of a particular “best” method may result in too few observations, in which case it may be more fruitful to select the most frequently used analytical method in the database. Data may have been recorded using analytical methods under separate synonymous names, or analytical methods incorrectly entered when data were first added to the database. Review of recorded data and analytical methods recorded by knowledgeable personnel is necessary to correct these problems.

Laboratory Quality Control

Laboratory quality control of data (blanks, spikes, replicates, known standards, etc.) where available should be reported. Such information may have been infrequently reported in larger data repositories and needs to be identified and coded. Records of general laboratory quality control protocols and specific quality control procedures associated with specific datasets are valuable in evaluating data quality. However, premature elimination of lower quality data can be counterproductive because the increase in variance caused by analytical laboratory error may be negligible compared with natural variability or sampling error, especially for nutrients and related water quality parameters. However, data of uncertain quality should not be accepted unless no other data are available.

Data Collecting Agencies

Selecting data from particular agencies with known, consistent sampling and analytical methods will reduce variability caused by unknown quality problems. Requesting data review for quality assurance from the collecting agency will reduce uncertainty about data quality.

Time Period

Long-term records are critically important for establishing trends. Determining if trends exist in the time series database is also important for characterizing reference conditions for nutrient criteria. Length of time series data needed for analyzing nutrient data trends is discussed in the Sampling Design section (Section 5.3).

Index Period

The index period for estimating average concentrations can be established if nutrient and water quality variables were measured through seasonal cycles. The index period may be the entire year or the summer season. The best index period is determined by considering water quality characteristics for the region, the quality and quantity of data available, and estimates of temporal variability (if available). Additional information and considerations for establishing an index period are discussed in Section 5.3.

Representativeness

Data may have been collected for specific purposes. Data collected for toxicity analyses, effluent limit determinations, or other pollution problems may not be useful for developing nutrient criteria. Furthermore, data collected for specific purposes may not be representative of the spatial scale of interest. The investigator must determine if the spatial scale for the data included in the database is representative of the area to be characterized. If a sufficient amount of data for the appropriate scale cannot be found, then new surveys will be necessary (see Section 5.3).

Gathering New Data

New data should be gathered following the sampling design protocols discussed below. New data collection activities for developing nutrient criteria should focus on filling in gaps where data are particularly needed for high-risk systems. Data gathered under new monitoring programs should be imported into databases or spreadsheets and merged with the existing nutrient database for criteria development.

5.3 SAMPLING DESIGN

This section discusses issues surrounding sampling nutrients, response variables, and related environmental variables in estuaries and coastal waters. Where appropriate data are unavailable or insufficient to derive numerical nutrient criteria, efforts must be made to collect new data to fill those gaps. New sampling programs should be scientifically based and statistically rigorous while maximizing available management resources. Such programs are used to better define nutrient and algal relationships within an ecosystem framework. At the broadest level, sampling efforts should detect or contribute to the following objectives:

- Identify the reference condition, that is, existing, most natural, least culturally impacted locations and their relative enrichment status

- Identify whether nutrient concentrations or loads are increasing, decreasing, or remaining the same

- Characterize seasonal patterns in nutrient levels and their relationship to primary productivity

- Help assess the assimilative capacity of the system, that is, contribute to the determination of how much nutrient loading can be assimilated without causing unacceptable changes in water quality or algal biomass and composition. (Note: In estuaries and coastal waters, this objective will likely require application of a computer-based hydrodynamic and nutrient-coupled water quality model. The intention here is to recognize particularly susceptible waters, not to lower expectations relative to historical antecedents and the reference condition.)

Some sampling programs may be poorly and inconsistently funded or are improperly designed and carried out, making it difficult to collect a sufficient number of samples over time and space to identify changes in water quality or to estimate average conditions with statistical rigor. This section provides a procedural approach for assessing water quality condition and identifying impairment by nutrients and algae in estuaries and coastal waters. The approaches described below present sampling designs that allow one to obtain a significant amount of information while attempting to minimize overall effort (and cost). Probabilistic and stratified random sampling begin with large-scale random monitoring designs that are reduced as nutrient and algal conditions are characterized. The tiered approach to sampling begins with coarse screening and proceeds to more detailed protocols as impaired and high-risk systems are identified and targeted for further investigation.

Sampling Protocol

Success of nutrient criteria development requires that consideration be given to sampling design. Initially, the relationship between critical response variables and nutrient concentrations, or in some cases, nutrient loads, needs to be established. Next, reference sites should be sampled, if feasible, in an attempt to establish reference conditions within classes of systems or subsystems. Classification should be linked to the reference condition activity. Nutrient concentrations/load and algal biomass relationships should help define the ecological state that can be attained if impaired systems are restored. As discussed in the following sections, this is not a straightforward exercise; it is very difficult to predict water transport/mixing in estuaries and coastal waters. The physics of these waters plays a major role in determining the observed patterns in nutrient and chlorophyll *a* concentrations, turbidity, and bottom water dissolved oxygen deficiency as well as transport. Variability in time and space further complicates empirical analysis as pointed out in Chapter 2.

Nutrient concentrations, chlorophyll *a* production, and system respiration represented as biological oxygen demand (BOD) are biochemical processes. How these processes are expressed in terms of dissolved oxygen, especially in bottom waters, has a great deal to do with the variability in freshwater flows, density stratification, advection, and mixing. The forcing factors include wind setups, changes in barometric pressure gradients, freshwater gravitational circulation, and the added complexity of bottom bathymetry. Interactions of these factors may create “flow jets,” flow reversals, three-layered circulation, and other physical complexities that suggest any monitoring scheme planned for estuaries and coastal waters would be advised to have a physical oceanographer as part of the team.

Sampling Technique

A reasonable and representative method is to profile the general physical character of the site by a CTD hydrocast. Water samples then may be collected from the surface 1 meter, mid-depth, and bottom 1 meter of the water column. Sample station activities should be coordinated as much as possible with the same tidal and current phase each time data are collected. If turbidity is measured by Secchi depth, the disc should be lowered from the shaded side of the vessel and depth determined from as close to the water level as practicable. Secchi depth measurements should be made only during periods of full daylight. The data for each station and sampling event should be recorded for each depth interval. This permits assessment for surface as well as bottom conditions. Where satisfactory, the results for each sampling episode can be combined into a mean or median measure representing all depths at that site. Temporal and spatial medians of the sites then can be determined to establish the representative values for that reference site.

During sampling visits, the candidate reference stations also should be examined to confirm whether they actually meet the reference site requirements. This may include looking for nearby discharges into the waters or tributaries and a quick survey of the shoreline to determine if new modifications may have changed the site. If an area appears to have been significantly impacted, measurements should be made for nutrient concentrations and biological response variables such as chlorophyll concentration and fish, macroinvertebrate, macrophyte, and planktonic community variables. Sites that do not meet the physical reference requirements should be excluded from the reference dataset. However, a high nutrient concentration present in an otherwise minimally developed area is not justification alone for exclusion. This may be part of the natural background level to be identified by the reference condition process.

Initial Considerations

Variability is inherent in sampling, which means that accuracy (how well the measure reflects actual conditions) and precision (how consistent the measurement is) must be assessed. Precision in ecological samples and measurements is more easily characterized than accuracy. Replicate samples from an experimental unit provide the basis for precision analysis. Standard statistical textbooks focus on precision (i.e., various ways to assess the nature of variability) and especially inferences regarding null hypotheses. In analytical chemical analyses, accuracy can be assessed by including samples of known purity and/or amount. Accuracy often refers to systematic errors in a method, whereas precision refers more or less to random errors. Outliers can be detected with statistical methods, but the so-called outlier may actually prove to be more accurate than the remainder of the data.

Some key questions to consider are the following: Does the method reflect more on how the assessor or analyst uses the method if the technician is imprecise, and Does the method actually do what it is purported to do? For example, is the arsenic concentration high enough that it interferes importantly with the phosphate analysis? Does the Loran system accurately place one on station, especially when electronic interferences are highly probable? Does the temperature setting on the dissolved organic carbon analyzer allow for accurate measurement or a biased one, or does the chlorophyll *a* method accurately correct for interferences from other pigments? For these examples, the precision may be very high (low variance) but the measurement may be in substantial error so the results are inaccurate. The following discussion expands on these ideas and provides additional information on statistical concepts and procedures relative to sample design. It is worth remembering that environmental data may not conform to the assumptions of normality required for statistical inferences; adjustments often can satisfy parametric assumption, but when they do not, the analysts must resort to distribution-free methods.

Specifying the Population and Sample Unit

Sampling is statistically expressed as a sample from a population of objects. Finite populations may be sampled with corresponding natural sample units, but often the sample unit (say, an estuary) is too large to measure in its entirety, and it must be characterized with one or more second-stage samples of the sampling gear (bottles, benthic grabs, quadrats, etc.). Each sample unit is assumed to be independent of other sample units. The objective of sampling is to best characterize individual sample units in order to estimate some attributes (e.g., nutrient concentrations or dissolved oxygen) and their statistical parameters (e.g., mean, median, variance and percentiles) of a population of sample units. The objective of the analysis is to be able to say something (estimate) about the population. It is critical to distinguish between making an inference about a population of many estuaries (e.g., “lagoonal estuaries around northern Gulf of Mexico are shallow and mesotrophic”) versus an inference about a single estuary or coastal water (e.g., “estuary ABC has fewer fish species than unimpaired reference estuaries or salinity zones within estuaries”). These two kinds of inferences require different sampling designs: the first requires independent observations of many waterbodies and does not require repeated observations within sample units (pseudoreplication; Hurlbert 1984), while the second often does require repeated observations within a waterbody. Examples of sample units include:

- A point in an estuary or coastal water (may be characterized by single or multiple sample device deployments). The population then would be all points in the waterbody, an infinite population.

- A constant area (e.g., square meter, hectare). The population could be all square meters of a coastal water surface area in a State or region.

- An estuary or a definable subbasin or salinity zone of an estuary as a single unit. Because salinity often specifies population distributions in estuaries, these zones most often are discrete environments, at least in the short term, and this is likely to be the most common sample unit. The population would be all salinity zones in a State or region, a finite population.

Specifying the Reporting Unit

It is also necessary to specify the units for which results will be reported. Usually, this is the population (e.g., all estuaries), but it also can be subpopulations (e.g., estuaries within a given nutrient ecoregion) and even individual locations (e.g., estuaries or coastal waters of special interest). To help develop the sampling plan, it is useful to create hypothetical statements of results in the way that they will be reported, for example:

- Status of a place: "The estuary ABC is degraded."

- Status of a region: "An estimated 20% of the estuary area in State XYZ has an elevated trophic state, above reference expectations"; "Approximately 20% of estuaries in State XYZ have an elevated trophic state."

- Trends at a place: "Nutrient concentrations in estuary ABC have decreased by 20% since 1980."

- Trends of a region: "Average estuary trophic state in State XYZ has increased by 20% since 1980"; "Average trophic state index values in 20% of estuaries of State XYZ have increased by 15% or more since 1980."
- Relationships among variables: "A 50% increase of N loading above natural background is associated with decline in taxa richness of benthic macroinvertebrates, below reference expectations."; "Coastal waters receiving runoff from large nonpoint sources have 50% greater probability of elevated trophic state above reference conditions than coastal waters not receiving such runoff."

Sources of Variability

Variability of measurements has many possible sources, and the intent of many sampling designs is to minimize the variability due to uncontrolled or random effects, and conversely to be able to characterize the variability caused by experimental or class effects. For example, estuaries may be classified by soil phosporus content of their surrounding watersheds so that estuaries within a class are likely to have similar water column concentrations in current or historical reference areas. The population of estuaries is stratified so that observations (sample units) from the same stratum will be more similar to each other than to sample units in other strata.

Environmental measures vary across different scales of space and time, and sampling design must consider the scales of variation. In coastal waters, measurements of some variables such as total nitrogen or chlorophyll concentrations are taken at single points in space and time (center of the deep depression, 20 m depth, 10 a.m. on 2 July). If the same measurement is taken at a different place (littoral zone, 1 m), or coastal waters, or time (30 January), the measured value may be different. A third component of variability is the ability to accurately measure the quantity of interest, which can be affected by sampling gear, instrumentation, errors in proper adherence to field and laboratory protocols, and the choice of methods used in making determinations.

The basic rule of efficient sampling and measurement is to sample so as to minimize measurement errors, to maximize the components of variability that have influence on the central questions and reporting units, and to control other sources of variability that are not of interest, that is, to minimize their effects on the observations. In the example of chlorophyll concentrations, variability could be reduced by sampling each of several coastal waters in the deepest part, with multiple depth samples or a vertically integrated pump sample taken in early spring before stratification appears. Many coastal waters are sampled to examine and characterize the variability due to different coastal waters (the sampling unit). Each coastal water is sampled in the same way, in the same place, and in the same timeframe to minimize variability due to location, depth, and season, which are not of interest in this particular study.

In the above example, chlorophyll concentrations vary with location within a coastal water, among coastal waters, and time of sampling (day, season, year). If the spatial and temporal components of variability within coastal waters are large (e.g., measurements of chlorophyll concentrations typically vary more between spring and fall samples within a coastal water than they do among coastal waters), then it may be best to use an index period. For this reason, coastal water chlorophyll concentrations often are estimated as a growing season average, estimated from several determinations (e.g., monthly) during the growing season.

In statistical terminology, there is a distinction between sampling error and measurement error that has little to do with actual errors in measurement. Sampling error is the error attributable to selecting a certain sample unit (e.g., a coastal water or a location within a coastal water) that may not be representative of the population of sample units. Statistical measurement error is the ability of the investigator to accurately characterize the sampling unit. Thus, measurement error includes components of natural spatial and temporal variability within the sample unit as well as actual errors of omission or commission by the investigator. Measurement error is minimized with methodological standardization: selection of cost-effective, low-variability sampling methods; proper training of personnel; and quality assurance procedures to minimize methodological errors. In analytical laboratory procedures, measurement error is estimated by replicate determinations on some subset of samples (but not necessarily all). Similarly, in field investigations, some subset of sample units should be measured more than once to estimate measurement error.

Analysis of variance (ANOVA) can be used to estimate measurement error. All multiple observations of a variable are used (from all coastal waters with multiple observations), and coastal waters are the primary effect variable. The root means square error (RMSE) of the ANOVA is the estimated variance of repeated observations within coastal waters. Note that a hypothesis test (F-test) is not of interest in this application, only the RMSE of the analysis.

Natural variability that is not of interest for the questions being asked, but which may affect the ability to address them, should be estimated with the RMSE method above. If the variance estimated from RMSE is unacceptably large (i.e., as large or larger than variance expected among sample units), then it is often necessary to alter the sampling protocol, usually by increasing sampling effort in some way to further reduce the measurement error. Measurement error can be reduced by multiple observations at each sample unit, for example, multiple ponar casts at each sampling event; multiple observations in time during a growing season or index period; depth-integrated samples; or spatially integrated samples.

A less costly alternative to multiple measures in space is spatially composite determinations. In nutrient or chlorophyll determinations, a water column pumped sample, where the pump hose is lowered through the water column, is an example of a spatially composite determination. Spatial integration of an observation and compositing the material into a single sample is almost always more cost-effective than retaining separate, multiple observations. This is especially so for relatively costly laboratory analyses such as organic contaminants and benthic macroinvertebrates, but the price of this economy is loss of information about the water column or about distribution over an area.

Statistical power is the ability of a given hypothesis test to detect an effect that actually exists and must be considered when designing a sampling program (e.g., Peterman 1990, Fairweather 1991). The power of a test (1-b) is defined as the probability of correctly rejecting the null hypothesis when the null hypothesis is false (i.e., the probability of correctly finding a difference [impairment] when one exists). For a fixed confidence level (e.g., 90%), power can be increased by increasing the sample size or the number of replicates, except in cases where the variance is proportional to the mean. To evaluate power and determine sampling effort, an ecologically meaningful amount of change in a variable must be set.

Optimizing sampling design requires consideration of tradeoffs among the measures used, the effect size that is considered meaningful, desired power, desired confidence, and resources available for the sampling program. Every study requires some level of repeated measurement of sampling units to estimate precision and measurement error. Repeated measurement at 10% of sites is common among many monitoring programs.

Alternative Sampling Designs

Sampling design is the selection of a part of a population to observe the attributes of interest to estimate the values of those attributes for the whole population. Classical sampling design makes assumptions about the variables of interest; in particular, it assumes that the values are fixed (but unknown) for each member of the population until that member is observed (Thompson 1992). This assumption is perfectly reasonable for some variables, say, length, weight, and sex of members of an animal population, but it seems less reasonable for more dynamic variables such as nutrient concentrations, loadings, or chlorophyll concentrations of estuaries. Designs that assume that the observed variables are themselves random variables are model-based designs, where prior knowledge or assumptions are used to select sample units.

Probability-Based Designs (Random Sampling)

The most basic probability-based design is simple random sampling, where all possible sample units in the population have the same probability of being selected; that is, all possible combinations of n sample units have equal probability of selection from among the N units in the population. If the population N is finite and not excessively large, a list can be made of the N units, and a sample of n units is randomly selected from the list. This is termed list frame sampling. If the population is very large or infinite (such as locations in an estuary), one can select a set of n random (x,y) coordinates for the sample.

All sample combinations are equally likely in simple random sampling, thus there is no assurance that the sample actually selected will be representative of the population. Other unbiased sampling designs that attempt to acquire a more representative sample include stratified, systematic, multistage, and adaptive designs. In stratified sampling, the population is subdivided or partitioned into strata, and each stratum is sampled separately. Partitioning is typically done so as to make each stratum more homogeneous than the overall population; for example, estuaries could be stratified on ecoregion or coastal waters by dominant current structure. Systematic sampling is the systematic selection of every k^{th} unit of the population from one or more randomly selected starting units, and it ensures that samples are not clumped in one region of the sample space. Multistage sampling requires selection of a sample of primary units, such as fields or hydrologic units, and then selection of secondary sample units, such as plots or estuaries within each primary unit in the first-stage sample.

Estimation of statistical parameters requires weighting of the data with inclusion probabilities (the probability that a given unit of the population will be in the sample) specified in the sampling design. In simple random sampling, inclusion probabilities are by definition equal, and no corrections are necessary. Stratified sampling requires weighting by the inclusion probabilities of each stratum. Unbiased estimators have been developed for specific sampling designs and can be found in sampling textbooks, such as Thompson (1992).

Model or Goal-Based Designs

Use of probability-based sampling designs may miss relationships among variables (models), especially if there is a regression-type relationship between an explanatory and a response variable. As an example, elucidation of estuary response to N loading with the Vollenweider-type model; that is, chlorophyll *a* concentration regressed against a depth-normalized N concentration (Vollenweider 1968) requires a range of trophic states from ultra-oligotrophic to hypereutrophic. A simple random sample of estuaries is not likely to capture the entire range (i.e., there would be a large cluster of "mesotrophic" estuaries with few at high or low ends of the trophic scale), and the random sample therefore may be biased with respect to the model.

In model-based designs, sites are selected based on prior knowledge of auxiliary variables, such as estimated phosphorus loading, estuary depth, and elevation. Often, these designs preclude an unbiased estimate of the population response variable (e.g., trophic state), unless the model can be demonstrated to be robust and predictive, in which case the population value is predicted from the model and from prior knowledge of the auxiliary (predictive) variables. Selection of unimpacted reference sites is an example of samples for a model (index development; response of index variables to measures of anthropogenic influence) that cannot later be used for unbiased estimation of the biological status of estuaries. Ideally, it may be possible to specify a design that allows unbiased estimation of both population and model. Statisticians should be consulted in developing the sample design for a nutrient criteria program.

Sampling and Analytical Designs for More Complex Ecological Questions

Complex ecological questions may not be required to develop numerical nutrient criteria. However, the manager may require a biological and an ecological assessment of resources at risk to establish management goals—that is, focus on biological resources of high social and economic value (e.g., the Chesapeake Bay Program includes biological variables as part of the goal setting process). Questions on how to sample different levels of biological organization (e.g., populations, communities, and ecosystems), indicators of stress, diversity and similarity measures, and biotic indices may become important. Criteria that go beyond the core variables will likely address one or more of these ecosystem or community elements. Publications are available that provide conceptual and statistical guidance for monitoring biological/ecological systems, including multivariate analytical approaches (e.g., Spellerberg 1991, Luepke 1979, Digby 1987, Clark and Warwick 1994a,b, Ott 1995, Ludwig and Reynolds 1988, Eckblad 1991).

Monitoring Programs

The purpose of monitoring is to obtain data that can be used not only to determine reference condition, but to help classify estuaries and coastal waters, or portions thereof, into groups (see Chapter 3). Classification should aid in the determination of reference sites or stations that are representative and have the lowest possible variability.

In some cases, a problem may exist where monitoring data indicate that the system has been greatly impaired from nutrient enrichment over the period of record. This is analogous to the so-called "corn-belt" problem in some lakes in the upper mid-West of the United States (see Lakes and Reservoirs Nutrient Guidance Document). This problem suggests that a meaningful reference condition may no longer exist. Or, the system has been greatly disturbed and it is not clear to what extent the impairments are due to nutrient enrichment. In this case, both historical information and diagnostic sampling may be required to clarify the reference condition and subsequent nutrient criteria.

Where data are insufficient, several approaches can be tried, for example, running a mathematical nutrient model "backwards"; use of biostratigraphic approaches, including changes in algal dominance and composition, loss of submerged aquatic vegetation (SAV), and pyritization of iron in sediments to detect earlier anoxia (Brush 1984, 1986, 1992, Cooper 1995) or reference to old written accounts (e.g., newspapers, diaries). For example, there are accounts of water clarity in the mouth of the Patuxent River estuary, Chesapeake Bay, in the late 1930's where engineers sitting in a "Beebe-like Bathysphere" on the estuary bottom could see horizontally approximately 20 to 30 feet. The methods discussed here are often qualitative to semiquantitative, but such information can be useful, especially if marked nutrient increases have evidently occurred over historical conditions but ambient data are insufficient. Older aerial photographs and other forms of watershed land use information and human population density trends can help make extrapolations regarding the system's response to nutrient loading.

Parameters to Survey

Each of the core variables discussed in Chapter 4 must be included in the survey (e.g., concentrations of TN, TP [total phosphours], chlorophyll *a*, and a measure of water clarity, such as Secchi disc, submersible PAR meter, or spectral radiometer). It is also appropriate to measure salinity, water temperature, flow and direction, tide phase, pH, and nutrient load to help better interpret the core variables. This is a much different problem than usually experienced in rivers and lakes. It may be desirable in some circumstances to include secondary variables, for example, vertical dissolved oxygen profiles, distribution and abundance of SAV/seagrasses, distribution of tidal emergent marshes, distribution and density of benthic filter feeders (e.g., oysters), water color, dissolved organic carbon (especially if humic-like materials are abundant), and particulate organic carbon. This more complex array of variables would require a diagnostic justification.

Sampling Frequency

A single grab sample from an estuary or coastal water will be grossly inadequate. Estuaries are near the bottom of watersheds, which makes them prone to episodic rainfall events. Coastal waters are also subject to seasonal storms that churn the waters and physically disturb shallow sediments, and these events may be seasonally highly variable. If information is available to set expectations when possible seasonal pulses of freshwater occur, then it should be used to help schedule the sampling of wet and dry periods. In north temperate estuaries, where winter to early spring is the dominant freshet period, this interval should be included in the sampling scheme. A lag of hours, days, or several weeks to one or more months is usually required to detect the system’s response to the nutrient load, depending on magnitude of the freshet relative to volume of the estuary or mixing zones of coastal waters. This also may capture any spring blooms of diatoms if such occur. A midsummer and early fall survey should give a first-order picture of the nutrient concentration and response variable pattern suitable for classification. In the event of variable summer freshwater flows, then more frequent sampling may be required. Because different patterns of rainfall exist around the coasts, regional considerations should weigh heavily in the design of sampling schedules.

Long-term datasets have well-documented ecological value (Likens 1992, Wolfe et al. 1987, Livingston 2001a); however, all too frequently resources constrain longer term sampling which can average out short-term variability. Recent data connected to long-term trends provide the strongest case for classification, reference condition determination, and other criteria development. By measuring the nutrient load, especially during freshet and low flow periods and concurrently with ambient water quality and hydrographic sampling, one can get an estimate of the load and salinity/nutrient and response variable relationships while keeping in mind the precautions noted above. For comparative purposes, it is important to compare core monitoring variables under similar salinity conditions.

If tidal elevation is large (e.g., greater than 2.0 m), then this component of estuarine flushing probably dominates over nontidal gravitational flows (Monbet 1992), and eutrophication symptoms are likely to be of a small magnitude. In some estuaries (e.g., York River estuary of the Chesapeake Bay), spring tides may break down density stratification, and the system responds differently to nutrient supplies than during periods of relatively strong stratification (Hass 1977). Thus, for estuaries with tidal elevations less than 2.0 m, it is important to note that they are likely to be quite vulnerable to nutrient enrichment.

A general rule of thumb regarding freshwater run-off events to estuaries is that a large freshet may displace the nutrient supply and responses will be detected seaward of the focal area. A modest freshet may not deliver enough nutrients or physically affect the density stratification to make an estuary vulnerable to nutrient enrichment. But an intermediate freshet may cause the focal area to receive a significant nutrient load and establish a strong vertical density gradient so maximum responses will be detected (e.g., high average chlorophyll *a* concentrations and minimum Secchi disk readings). This rule is less easily applied to coastal waters.

Sampling Locations

Sampling locations depend on the size (and especially the length of an estuary), bathymetry, nutrient source inputs, and hydrography (especially the longitudinal and vertical salinity profiles). In estuaries, consideration should be given to tidal freshwater, the turbidity maximum (if one is present), mesohaline and polyhaline regimes, as well as water below zones of density stratification.

In large tidal freshwater riverine systems, it is important to employ several stations because this portion of the estuary may "store" a large supply of nutrients that later advect into the saline reach of the estuary (e.g., the Hudson River system) (Lampman et al. 1999). Enough samples should be taken to detect nutrient concentration gradients along the salinity gradient from tidal river to the estuary receiving waters. Typically, this will require from five to seven stations at a minimum. If the estuary is relatively wide (e.g., lagoonal systems such as Pamlico Sound, NC, or Pensacola Bay, FL) or has large tributary creeks, then these features may need independent sampling. Where salinity gradients are distinct both horizontally and vertically, composite sampling may have severe limitations. Depth variability also should be considered, for example, main channel, shelf samples, and samples in shallow water near SAV/seagrass meadows or in emergent marsh channels should be included. Emergent marsh creeks should be sampled in the summer during high and low tides, because high system respiration may cause hypoxia/anoxia in these tidal creeks that may be largely natural. Where SAV meadows are poorly developed, resuspension of bottom sediments may be more common and not represented by open channel samples.

Serious consideration should be given to some replicate sampling within salinity zones to estimate variability; however, resources may require a broad picture where gradients become equally or more important than the physical salinity "zones." In most cases, analytical levels of detection should be a trivial aspect of data acquisition for reference characterization. This does not free one from application of good laboratory quality asurance/quality control (QA/QC) practices, which must be maintained with appropriate blanks, reference samples, and other considerations to standard analytical measurements.

Citizen Monitoring Programs

Citizen monitoring programs have greatly increased, especially since the early 1980's. Where there is adequate technical oversight either from within the group expertise or from the outside, such monitoring efforts can play an important role in assessing trends, identifying "hotspots," and locating likely sources of nutrients, especially in smaller estuaries where larger research vessels are not required. Many Federal, State, Tribal, and local agencies assist citizen monitoring efforts, and these agencies contribute to training and direction, development, and implementation of QA/QC procedures, act as a data repository; and perform analyses on environmental samples collected by citizen groups.   Citizen monitoring groups often can provide more frequent observations, such as visiting a gauging station, than can State personnel. Citizens also can identify those property holders or resource users not following best management practices or operating within permit limits. See also Volunteer Monitoring Programs, above.

5.4 QUALITY ASSURANCE/QUALITY CONTROL

The validity and usefulness of data depend on the care with which they were collected, analyzed, and documented. EPA provides guidance on data QA/QC (U.S. EPA 1998b) to assure the quality of data. Factors that should be addressed in a QA/QC plan are briefly described below, but the reader is referred to published EPA guidance for specifics. The QA/QC plan should state specific goals for each factor and should describe the methods and protocols used to achieve the goals. The five factors discussed below are representativeness, completeness, comparability, accuracy, and precision.

Representativeness

Sampling program design (when, where, and how you sample) should produce samples that are representative or typical of the environment being described. Sampling designs for developing nutrient criteria are discussed earlier in this chapter.

Completeness

Datasets are often incomplete because of spilled samples, faulty equipment, and/or lost field notebooks. A QA/QC plan should describe how complete the dataset must be to answer the questions posed (with a statistical test of given power and confidence) and the precautions being taken to ensure that completeness. Data collection procedures should document the extent to which these conditions have been met. Incomplete datasets may not invalidate the collected data, but they may reduce the rigor of statistical analyses. Therefore, precautions should be taken to ensure data completeness. These precautions may include collecting extra samples, having backup equipment in the field, installing alarms on freezers, copying field notebooks after each trip, and/or maintaining duplicate sets of data in two locations.

Comparability

To compare data collected under different sampling programs or by different agencies, sampling protocols and analytical methods must demonstrate comparable data. The most efficient way to produce comparable data is to use sampling designs and analytical methods that are widely used and accepted such as Standard Methods for the Examination of Water and Wastewater (APHA, AWWA, WEF, 1998) and EPA methods manuals.

Accuracy

To assess the accuracy of field instruments and analytical equipment, a standard (a sample with a known value) must be analyzed and the measurement error or bias determined. Internal standards should periodically be checked with external standards provided by acknowledged sources. At Federal, State, Tribal, and local government levels, the National Institute of Standards and Technology (NIST) provides advisory and research services to all agencies by developing, producing, and distributing standard reference materials. For calibration services and standards see: http://ts.nist.gov/ts/htdocs/230/233/home/calibration.html.

Standards and methods of calibration are typically included with CTD sondes, turbidity meters, pH meters, DO meters, and DO testing kits. USGS, EPA, and some private companies provide reference standards or QC samples for nutrients. Reference standards for chlorophyll are also available from EPA and some private companies, although chlorophyll standards are time and temperature sensitive because they degrade over time.

Variability

Natural variability, rather than imprecision in the method used, is usually the greatest source of error in the constituent measured. The variability in field measurements and analytical methods should be demonstrated and documented to identify the source of variability when possible. EPA QA/QC guidance provides an explanation and protocols for measuring sampling variability (U.S. EPA 1998). Methods for creating a chlorophyll standard to determine if the spectrophotometer is measuring chlorophyll consistently from one year to the next or from the beginning to the end of an analytical run are described in Wetzel and Likens (1991). In addition, replicates for each sample time and site (usually three) must be collected because the largest source of variation is likely to be natural (i.e., in the samples).

5.5 STATISTICAL ANALYSES

Statistical analyses are used to identify variability in data and to elucidate relationships among sampling parameters. Several statistical approaches for analyzing data are mentioned here. We advocate simple descriptive statistics for initial data analyses, that is, calculating the mean, median, mode, ranges, and standard deviation for each parameter in the system of interest. The National Nutrients Database discussed above calculates simple descriptive statistics for queried data. Specific recommendations for setting criteria using frequency distributions are discussed in Chapter 7.

Data Reduction

Data reduction requires a clear idea of the analysis that will be performed and a clear definition of the sample unit for the analysis. For example, a sample unit might be defined as “an estuary during July–August.” For each variable measured, a median value then would be estimated for each estuary in each July–August index period on record. Analyses then are done with the observations (estimated medians) for each sample unit, not with the raw data. Steps in reducing the data include:

- Selecting the long-term time period for analysis
- Selecting an index period
- Selecting relevant chemical species
- Identifying the quality of analytical methods
- Identifying the quality of the data recorded
- Estimating values for analysis (mean, median, minimum, maximum) based on the reduction selected.

Frequency Distributions

Frequency distributions can be used to aid in the setting of criteria. Frequency distributions do not require prior knowledge of individual waterbody conditions before setting criteria. Criteria are based on and, in a sense, developed relative to the population of systems in the Region, State, or Tribal jurisdiction.

Data plotted on a scale of mean nutrient concentration versus frequency of occurrence for a specific estuary, portion of an estuary, or coastal reach produces a frequency distribution of mean or median nutrient concentration. Plots of frequency distributions of median TP, median TN, median chlorophyll *a*, and Secchi depth for the index period (discussed in Chapter 4) should be examined to determine the normality of the data in the distribution and to determine the potential for further subdivision of the waterbody under investigation. Data that are not normally distributed often are transformed into a distribution more approximating the normal distribution by taking the logarithm of each value. Analysis of outliers may assist in explaining variability in small datasets; additional analysis can be conducted to identify the statistical significance of population differences.

Correlation and Regression Analyses

The relationship between two variables may be of use in analyzing data for criteria derivation. Correlation and regression analyses allow the relationship to be defined in statistical terms. A correlation coefficient, usually identified as r, can be calculated to quantitatively express the relationship between two variables. The appropriate correlation coefficient is dependent on the scale of measurement in which each variable is expressed (whether the distribution of data is continuous or discrete) and whether there is a linear or nonlinear relationship. Results of correlation analyses may be represented by indicating the correlation coefficient and represented graphically as a scatter diagram that plots all of the collected data, not just a measure of central tendency. The statistical significance of a calculated correlation coefficient can be determined with the t test. The t test is used to determine if there is a true relationship between two variables. Therefore, the null hypothesis states that there is no correlation between the data variables measured within the population. A critical α value is chosen as a criterion for determining whether to reject the null hypothesis. If the null hypothesis is rejected, the alternate hypothesis states that the correlation at the calculated r value between the two variables is significant.

Regression analysis provides a means of defining a mathematical relationship between two variables that permits prediction of one variable if the value of the other variable is known. In contrast to correlation analysis, there should be a true independent variable (a variable under the control of the experimenter) in regression analysis. Regression analysis establishes a relationship between two variables that allows prediction of the dependent variable (predicted variable) for a given value of an independent variable (predictor variable). However, scientists (other than statisticians) apply regression analyses to field data when a relationship is known to exist, even when there is no true independent variable (e.g., cell counts of algae and chlorophyll concentration; nutrient concentrations and chlorophyll concentration) (Ott 1988, Campbell 1989, Atlas and Bartha 1993, Ott 1995).

Tests of Significance

Various statistical tests are used to assess the hypotheses being tested. Statistical tests of significance differ in their applicability to the dataset of interest and the power of the test (the ability of the test to detect a false null hypothesis). A parametric test of significance assumes a normal distribution of the population. Nonparametric analyses are valid for any type of distribution (normal, log-normal, etc.) and can be used if the data distribution is not normal or unknown. A parametric test has more power than a nonparametric test when its assumptions are satisfied. Two types of errors can be made when testing hypotheses: Type I—where a correct null hypothesis is mistakenly rejected, and Type II—when there is a failure to reject a false null hypothesis. The parametric test is less likely than a nonparametric test to make a Type II error, when the assumptions are met. Therefore, if given a choice, the parametric test should be used rather than the nonparametric test when the assumptions of the parametric test are fulfilled. Less powerful, nonparametric tests of significance must be used in cases where the data do not fit the assumption of a normal distribution (Ott 1988, Campbell 1989, Atlas and Bartha 1993). Parametric tests include the student *t* test, analysis of variance, multivariate analysis of variance, and multiple range tests. Nonparametric tests include chi square, Mann Whitney U test, and the Kruskal–Wallis test (Ott 1988; Campbell 1989; Atlas and Bartha 1993). Detailed descriptions of these and other relevant statistical tests can be found in standard statistical texts.